Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 435 – 454
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Estimating common vector parameters in interlaboratory studies

Andrew L. Rukhin[a, b, *]

[a] *Statistical Engineering Division, National Institute of Standards and Technology, Building 820, Gaithersburg, MD 20899, USA*

[b] *Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA*

Received 31 March 2005
Available online 30 October 2006

Abstract

The primary goal of this work is to extend two methods of random effects models to multiparameter situation. These methods comprise the DerSimonian–Laird estimator, stemming from meta-analysis, and the Mandel–Paule algorithm widely used in interlaboratory studies. The maximum likelihood estimators are also discussed. Two methods of assessing the uncertainty of these estimators are given.

AMS 1991 subject classification: Primary 62H12; secondary 62H10; 62P10

Keywords: Collaborative study; Consensus mean; DerSimonian–Laird estimator; Mandel–Paule algorithm; Meta-analysis; Random effects model; (Restricted) maximum likelihood estimator

1. Introduction and summary

Statistical modeling and analysis of collaborative studies pose several fundamental questions about determination of the consensus (reference) value and its associated uncertainty. An appropriate choice of stochastic model can be especially difficult especially when measurements are made across a range of values of a physical characteristic, i.e. the reference value is a curve or a multivariate vector. Hedges and Olkin [7] discuss several methods for combining data from several experiments which form the heart of the meta-analysis. In particular, in Chapter 10 these

* Corresponding author at: Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Fax: +1 410 455 1066.

E-mail address: rukhin@math.umbc.edu.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.09.005

authors discuss multivariate models for effect size when observations from experimental and control group are available.

Multivariate data sets occur increasingly often in applications in interlaboratory metrological studies known as Key Comparisons when one has to determine the key comparison reference value (KCRV), i.e. the consensus vector mean. In this paper, statistical procedures are derived for consensus vector evaluation (a discretized version of possibly irregular underlying curve) along with estimates of the uncertainty of this value. An approach in spirit of meta-analysis for interlaboratory data is proposed when participating laboratories exhibit different accuracy. This model for Gaussian distributions leads to a class of matrix-weighted vector statistics to estimate the common vector mean and to a method of assessing the uncertainty of these estimates. In Section 2, maximum likelihood estimators are reviewed including the restricted maximum likelihood estimator. Since these estimators do not admit explicit form, and the likelihood equations may have several roots, we suggest to employ in practical situations simpler procedures, scalar versions of which are widely used in biological and physical applications and which deserve more attention from statistical community. These scalar procedures are the DerSimonian–Laird estimator and the Mandel–Paule algorithm. Their method of moments origin and their relationship to the maximum likelihood estimator and to the restricted maximum likelihood estimator are discussed in Section 3.

We investigate the estimation problem of the covariance matrix of these procedures in Section 4 where approximate confidence ellipsoids are constructed. Section 5 contains results of Monte-Carlo simulation study and answers for motivating examples for this study. One of these examples is the Key Comparisons of accelerometers (CCAUV.V-K1) [11] which was organized to compare measurements of sinusoidal linear accelerometers over the range of frequencies from 40 Hz to 5 kHz. It was the first such study in the field of vibration and shock with the task of measuring the charge sensitivity of two accelerometers standards (we report results only for single-ended design accelerometer). Each of the 12 participating National Metrology Institutes measured charge sensitivity at the specified frequencies by employing two transfer standard accelerometers under agreed physical conditions. The common charge sensitivity and its uncertainty was one of the goals of the study. The participating institutes reported their means and the sample covariance matrices for the frequencies in this range. We suggest two methods for estimating the reference value for charge sensitivity as a function of frequency and give confidence ellipsoids. In the original study this value is found separately for each type of accelerometer and for each specified frequency.

Another example is the study of Pyroceram 9606, a glass ceramic material especially suited for high temperature applications. This material is being used for performance evaluation of instruments measuring thermal properties such as thermal conductivity, thermal diffusivity, and specific heat (heat capacity). All these characteristics are temperature dependent, so the reference value must be a function of temperature. Twenty-eight thermal conductivity experiments in different countries have been performed on this material, and a consensus value for diffusivity and heat was needed. We have used unpublished data (participating laboratories results were of widely differing quality) kindly provided by J. Filliben and R. Zarr (National Institute of Standards and Technology).

2. The matrix formulation of the model and likelihood equations

In accordance with the goals discussed in Section 1 we formulate the following mathematical model in the situation where multiple (correlated) q-dimensional measurements are made by each of p laboratories. In our model, the ith laboratory repeats its vector measurements $n_i (> q)$ times, and the vector data X_{ij} for $i = 1, \ldots, p$ and $j = 1, \ldots, n_i$ follow a one-way random effects MANOVA model, which may be both unbalanced and heteroscedastic, i.e.

$$X_{ij} = \theta + \ell_i + \varepsilon_{ij}, \tag{1}$$

with mutually independent $\ell_i \sim N_q(0, \Xi)$ and $\varepsilon_{ij} \sim N_q(0, \Delta_i)$, $j = 1, \ldots, n_i$. The vector θ plays the role of the common mean or the reference value, ℓ_i is the between-laboratories effect, and the ε's are the measurement errors. The unknown $q \times q$ matrix Ξ may have rank smaller than q; θ represents an unknown q-dimensional parameter common to all laboratories. The goal is to estimate the structural parametric vector θ, and to provide a standard error for this estimate. The covariance matrices Δ_i and Ξ are nuisance parameters.

In matrix notation this model can be written as a particular case of the general linear model

$$\mathbf{y} = \mathbf{T}\theta + \boldsymbol{\ell} + e.$$

Here $\mathbf{y}$ is the total data vector of dimension $n = q(n_1 + \cdots + n_p)$; $\mathbf{T}$ is a matrix of size $n \times q$ formed by $q \times q$ identity matrices written as a column; and n-dimensional vector $\boldsymbol{\ell}$ is formed by stacked n_i copies of $\ell_i, i = 1, \ldots, p$. Thus θ is the unknown parameter (fixed effects) vector, and $(\ell_1, \ldots, \ell_p)^T$ is random effect vector uncorrelated with the errors vector e.

The usual estimators of the laboratory means and of their covariance matrices are $X_i = \bar{X}_i = \sum_j X_{ij}/n_i$, and $S_i = \sum_j (X_{ij} - X_i)(X_{ij} - X_i)^T/[v_i n_i]$, $v_i = n_i - 1$, with $X_i, S_i, i = 1, \ldots, p$, being (incomplete) sufficient statistics. Reduction by sufficiency to the sample means X_i and the sample covariance matrices S_i makes this problem more specific. The model (1) leads to the following model for X_i,

$$X_i = \theta + \ell_i + \varepsilon_i. \tag{2}$$

Under the above assumptions, $\ell_i \sim N(0, \Xi)$, $\varepsilon_i \sim N(0, \Sigma_i = \Delta_i/n_i)$, with ℓ_i and ε_i being independent. Then, clearly,

$$Var(X_i) = \Xi + \Sigma_i.$$

The multiples of the sample covariance matrices, $v_i S_i$, are known to have a Wishart distribution $W_q(v_i, \Sigma_i)$.

2.1. The maximum likelihood estimators

The loglikelihood function ℓ for X_i and S_i can be written as

$$-2\ell = \sum_i [(X_i - \theta)^T(\Sigma_i + \Xi)^{-1}(X_i - \theta) + \log|\Sigma_i + \Xi|] + \sum_i v_i[\mathrm{tr}(S_i\Sigma_i^{-1}) + \log|\Sigma_i|].$$

The maximum likelihood estimator of θ has the form, $\hat{\theta} = \sum_{i=1}^p \hat{\omega}_i X_i$, where the matrix weights have the form

$$\hat{\omega}_i = \left[\sum_j (\hat{\Sigma}_j + \hat{\Xi})^{-1}\right]^{-1} \left(\hat{\Sigma}_i + \hat{\Xi}\right)^{-1},$$

and $\hat{\Sigma}_i$ and $\hat{\Xi}$ are found as maximizers of ℓ. When $q = 1$, this estimator was studied in [1,13,15].

The likelihood equations for Σ_i and Ξ are

$$\left(\hat{\Sigma}_i + \hat{\Xi}\right)^{-1}\left[\mathbf{I} - (X_i - \hat{\theta})(X_i - \hat{\theta})^T(\hat{\Sigma}_i + \hat{\Xi})^{-1}\right] + v_i \hat{\Sigma}_i^{-1}\left(\mathbf{I} - S_i \hat{\Sigma}_i^{-1}\right) = 0 \tag{3}$$

and

$$\sum_i \left[(\hat{\Sigma}_i + \hat{\Xi})^{-1}(X_i - \theta)(X_i - \theta)^T(\hat{\Sigma}_i + \hat{\Xi})^{-1} - (\hat{\Sigma}_i + \hat{\Xi})^{-1}\right] = 0. \tag{4}$$

Here and further $\mathbf{I}$ denotes the identity matrix whose dimension is clear from the context.

It follows that

$$\sum_i v_i \hat{\Sigma}_i^{-1}\left(\mathbf{I} - S_i \hat{\Sigma}_i^{-1}\right) = 0.$$

It is practical to determine the maximum likelihood estimators of θ, Σ_i's for fixed $\Xi = Y$. Thus, if for non-negative definite matrix Y

$$F(Y) = \min_{\theta, \{\Sigma_i\}} \sum_i \Big[(X_i - \theta)^T(\Sigma_i + Y)^{-1}(X_i - \theta) + \log|\Sigma_i + Y| \\ + v_i(\mathbf{tr}(S_i \Sigma_i^{-1}) + \log|\Sigma_i|)\Big],$$

then the maximum likelihood estimator of Ξ is $\arg\min F(Y)$.

The minimizers $\hat{\theta}(Y)$ and $\hat{\Sigma}_i(Y)$ can be determined from (3). An iterative scheme for solving (3) for fixed i and $\hat{\theta}$,

$$\Sigma_i^{(k+1)} = [v_i \mathbf{I} + B_k]^{-1}\,[v_i S_i + B_k(X_i - \hat{\theta})(X_i - \hat{\theta})^T B_k^T], \quad k = 0, 1, \ldots$$

with $B_k = \Sigma_i^{(k)}\left[\mathbf{I} + \Sigma_i^{(k)}\right]^{-1}$, converges fast. Eq. (4) shows that

$$F'(Y) = \sum_i v_i\left[\hat{\Sigma}_i^{-1}(Y) S_i \hat{\Sigma}_i^{-1}(Y) - \hat{\Sigma}_i^{-1}(Y)\right],$$

so that for any non-negative definite matrix Y_0,

$$F(Y) = F(Y_0) + \mathbf{tr}(F'(Y_0)(Y - Y_0)) + O(\|Y - Y_0\|^2).$$

The Hessian of F can be found from (3), but it has a complicated form. The constrained optimization problem (the matrix Y must be non-negative definite) is rather awkward. For these reasons, it is much more convenient to solve unconstrained optimization problems in $q \times r$-dimensional matrix $Z, r = 1, \ldots, q$ of rank r with $Y = Z^T Z$.

If $Y - Y_0 = V^T V$, then

$$(\hat{\Sigma}_i + Y)^{-1} = (\hat{\Sigma}_i + Y_0)^{-1} - (\hat{\Sigma}_i + Y_0)^{-1} V\left[\mathbf{I} + V^T(\hat{\Sigma}_i + Y_0)^{-1} V\right]^{-1} V^T(\hat{\Sigma}_i + Y_0)^{-1},$$

and

$$|\hat{\Sigma}_i + Y| = |\hat{\Sigma}_i + Y_0| \times |\mathbf{I} + V^T(\hat{\Sigma}_i + Y_0)^{-1} V|.$$

Therefore,

$$F(Y) \leqslant F(Y_0) - \sum_i \Big[(X_i - \hat\theta)^T(\hat\Sigma_i + Y_0)^{-1} V \left[\mathbf{I} + V^T(\hat\Sigma_i + Y_0)^{-1}V\right]^{-1} \times V^T(\hat\Sigma_i + Y_0)^{-1}(X_i - \hat\theta) - \log|\mathbf{I} + V^T(\hat\Sigma_i + Y_0)^{-1}V| \Big],$$

where $\hat\Sigma_i = \hat\Sigma_i(Y_0)$, and $\hat\theta = \hat\theta(Y_0)$. Therefore, one can construct a recursive sequence $Z_{k+1} = Z_k + t_k P_k, k = 0, 1, \ldots$ for $Y_k = Z_k^T Z_k$, in which the search directions U_k are taken to be the eigenvectors of $F'(Y_k)$ corresponding to negative eigenvalues and positive step sizes t_k are chosen as to minimize

$$\sum_i [\log|\mathbf{I} + t^2 U_k^T(\hat\Sigma_i + Y_k)^{-1}U_k| - t^2(X_i - \hat\theta_k)^T(\hat\Sigma_i + Y_k)^{-1} \times U_k[\mathbf{I} + t^2U_k^T(\hat\Sigma_i + Y_k)^{-1}U_k]^{-1}U_k^T(\hat\Sigma_i + Y_k)^{-1}(X_i - \hat\theta_k)].$$

Then, provided this minimum is attained at a positive t_k, $F(Z_{k+1}^T Z_{k+1}) \leqslant F(Z_k^T Z_k) = F(Y_k)$.

The estimator of $Var(\hat\theta)$ is usually obtained from the inverse of the observed Fisher information,

$$\widehat{Var}(\hat\theta) = \left[\sum_j (\hat\Sigma_j + \hat\Xi)^{-1}\right]^{-1}. \tag{5}$$

Sometimes the likelihood equations are solved via the EM-Algorithm in which the ℓ_i are interpreted as missing observations (see [14, Chapter 8]). This method works well when Σ_i do not depend on i, but for different Σ_i's leads to fairly cumbersome equations even when $q = 1$. Besides the recursive sequence obtained by this method is guaranteed to converge only to one of local maximums. The likelihood function may not be unimodal, i.e. can have several local extrema, so that all iterative algorithms are sensitive to the initial value. The estimator of Y discussed in Section 3.1 provides a good choice for this value.

In variance component problems the restricted maximum loglikelihood estimator is commonly recommended. To obtain the form of the restricted function, one can use the expression obtained in [5, p. 325]. It is based on the matrix formulation of the model (1) given in Section 2. The covariance matrix V of $\boldsymbol{\ell} + e$ is block-diagonal with $n_i q \times n_i q$ blocks formed by sums of diagonal blocks determined by the matrix Δ_i and blocks formed by Ξ. Therefore, V_i^{-1} has a similar form with the diagonal blocks Δ_i^{-1} and the off-diagonal entries $-\Delta_i^{-1}\Xi(\mathbf{I} + n_i\Delta_i^{-1})\Xi\Delta_i^{-1}$. Also $|V| = \prod_{i=1}^p |\Sigma_i|^{(n_i-1)}|n_i\Xi + \Delta_i|$.

Thus the restricted loglikelihood function ℓ_{R} in the variance parameters Σ_i and Ξ is to within an additive constant,

$$-2\ell_{\mathrm{R}} = \sum_i \left[(X_i - \hat\theta)^T(\Sigma_i + \Xi)^{-1}(X_i - \hat\theta) + \log|\Sigma_i + \Xi|\right] + \log\left|\sum_i (\Sigma_i + \Xi)^{-1}\right| + \sum_i v_i\left[\mathrm{tr}(S_i\Sigma_i^{-1}) + \log|\Sigma_i|\right]. \tag{6}$$

We will use (6) later in Section 3.2.

3. Weighted means statistics

Because of complicated nature of the maximum likelihood estimator, simpler procedures are desired. Even without normality assumption when the within-trials and between-trials covariance matrices Σ_i and Ξ are known, the least-squares estimator of the parameter θ in the model (1) is a weighted means statistic $\tilde{X}$. Let $W_i = [Var(X_i)]^{-1} = (\Sigma_i + \Xi)^{-1}$ and $\mathbf{W} = \sum_{i=1}^{p} W_i$. Then $\tilde{X}$ is found from the following equation,

$$\mathbf{W}\tilde{X} = \sum_{i=1}^{p} W_i X_i. \tag{7}$$

It makes sense to employ the available statistics S_i to approximate the within-trials covariance matrices Σ_i. In other terms we restrict the class of estimators (7) to those with matrix weights of the form

$$W_i = W_i(Y) = (S_i + Y)^{-1} \tag{8}$$

for some non-negative definite matrix Y. If $W^- = \left[\sum_{i=1}^{p} W_i(Y)\right]^-$, denotes a generalized inverse of $\sum_{i=1}^{p} W_i$, then an estimator $\tilde{X}$ of θ from this class has the following representation,

$$\tilde{X} = \tilde{X}_Y = W^- \sum_{i=1}^{p} W_i X_i = \sum_{i=1}^{p} \omega_i X_i. \tag{9}$$

Estimators of the form (9) are of interest. They include the analog of one of the traditional estimators of the common vector mean suggested by Graybill and Deal [4] in the case $q = 1$,

$$\tilde{X}_0 = \left[\sum_{i=1}^{p} S_i^{-1}\right]^- \sum_{i=1}^{p} S_i^{-1} X_i, \tag{10}$$

and the sample mean,

$$\tilde{X}_\infty = \frac{1}{p} \sum_{i=1}^{p} X_i.$$

Sometimes the within-laboratories variances Σ_i can be assumed to be known (in practice they are taken to be S_i). The maximum likelihood estimator $\tilde{X}_{\mathrm{SM}}$ of θ also is a weighted means statistic (9) with the weights of the form (8), where $Y = Y_{\mathrm{SM}}$ is the minimizer in Y of the negative loglikelihood function,

$$\sum_{1}^{p} \left[(X_i - \tilde{X})^T (Y + S_i)^{-1} (X_i - \tilde{X}) + \log |Y + S_i|\right]. \tag{11}$$

The resulting likelihood equation,

$$\sum_{1}^{p} \left[(Y + S_i)^{-1} (X_i - \tilde{X})(X_i - \tilde{X})^T (Y + S_i)^{-1} - (Y + S_i)^{-1}\right] = 0,$$

can be solved iteratively.

In the sequel we give two simpler methods of choosing the matrix Y.

3.1. DerSimonian–Laird procedure

If the matrix weights W_i are arbitrary, but the matrix $\mathbf{W}$ is non-singular, then for $\tilde{X}$ defined by (7),

$$\sum_i W_i^{1/2} E(X_i - \tilde{X})(X_i - \tilde{X})^T W_i^{1/2}$$
$$= \sum_{i=1}^{p} W_i^{1/2} \left(\mathbf{I} - \mathbf{W}^{-1} W_i\right) Var(X_i) \left(\mathbf{I} - \mathbf{W}^{-1} W_i\right)^T W_i^{1/2}$$
$$+ \sum_{i=1}^{p} W_i^{1/2} \mathbf{W}^{-1} \left(\sum_{k:k \neq i} W_k Var(X_k) W_k\right) \mathbf{W}^{-1} W_i^{1/2}. \tag{12}$$

In particular, when $W_i = \Sigma_i^{-1}$, $Var(X_i) = \Sigma_i + \Xi$,

$$\sum_i \Sigma_i^{-1/2} \left(\mathbf{I} - \mathbf{W}^{-1}\Sigma_i^{-1}\right) \Xi \left(\mathbf{I} - \mathbf{W}^{-1}\Sigma_i^{-1}\right)^T \Sigma_i^{-1/2}$$
$$+ \sum_i \Sigma_i^{-1/2} \mathbf{W}^{-1} \left(\sum_{k:k \neq i} \Sigma_k^{-1} \Xi \Sigma_k^{-1}\right) \mathbf{W}^{-1} \Sigma_i^{-1/2}$$
$$= \sum_i \Sigma_i^{-1/2} E(X_i - \tilde{X})(X_i - \tilde{X})^T \Sigma_i^{-1/2} - p\mathbf{I} + \sum_i \Sigma_i^{-1/2} \mathbf{W}^{-1} \Sigma_i^{-1/2}. \tag{13}$$

We suggest to use (13) as an estimating equation for the parameters θ and Ξ in the following way (provided that Σ_i are replaced by S_i). Let $\tilde{X}_0$ be the Graybill–Deal estimator (10). Put

$$\mathbf{B} = \sum_i S_i^{-1/2} (X_i - \tilde{X}_0)(X_i - \tilde{X}_0)^T S_i^{-1/2} - p\mathbf{I} + \sum_i S_i^{-1/2} \left(\sum_k S_k^{-1}\right)^{-1} S_i^{-1/2},$$

so that (symmetric) $\mathbf{B}$ estimates the right-hand side of (13). With $\widetilde{\omega}_i = [\sum_{j=1}^{p} S_j^{-1}]^{-1} S_i^{-1}$, determine a symmetric matrix Y from the equation

$$\sum_i S_i^{-1/2} \left(\mathbf{I} - \widetilde{\omega}_i\right) Y \left(\mathbf{I} - \widetilde{\omega}_i\right)^T S_i^{-1/2} + \sum_i S_i^{-1/2} \left(\sum_{j:j \neq i} \widetilde{\omega}_j Y \widetilde{\omega}_j^T\right) S_i^{-1/2} = \mathbf{B}. \tag{14}$$

Take $Y_{\mathrm{DL}} = Y_+$ to be the positive part of Y, i.e. let Y_{DL} have the same spectral decomposition as Y, with eigenvalues being positive parts of these of Y. The matrix weights of the estimator $\tilde{X}_{\mathrm{DL}}$ then are

$$W_i = (Y_{\mathrm{DL}} + S_i)^{-1}. \tag{15}$$

Eq. (14) extends the procedure suggested by DerSimonian and Laird [3] when $q = 1$, which is an immensely popular method in biostatistics especially in analysis of multicenter clinical trials. In fact, the number of references to the paper by DerSimonian and Laird exceeds 1500. This popularity is due mainly to the fact that this is a simple non-iterative procedure, which admits an approximate formula for the variance of the resulting estimator.

To solve (14), denote by $Vec(A)$ the $q^2 \times 1$ vector formed by stacking the columns of the $q \times q$ matrix A under each other, and by $A \bigotimes B$ the tensor (Kronecker) product of matrices A and B.

Then, according to Lemma 16.1.2 and Theorem 16.2.1 in [6]

$$
\begin{aligned}
&Vec\left(S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)\,Y\,(\mathbf{I}-\widetilde{\omega}_i)^T\;S_i^{-1/2}\right)\\
&\quad=\left[S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)\bigotimes S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)\right]Vec(Y),
\end{aligned}
$$

and

$$
Vec\left(\sum_i S_i^{-1/2}\left(\sum_{j:j\neq i}\widetilde{\omega}_j Y\widetilde{\omega}_j^T\right)S_i^{-1/2}\right)=\left(\sum_{i\neq j}S_i^{-1/2}\widetilde{\omega}_j\bigotimes S_i^{-1/2}\widetilde{\omega}_j\right)Vec(Y).
$$

Thus, the vectorized version of (14) holds when

$$
\begin{aligned}
&\left[\sum_i S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)\bigotimes S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)+\sum_{i\neq j}S_i^{-1/2}\widetilde{\omega}_j\bigotimes S_i^{-1/2}\widetilde{\omega}_j\right]Vec(Y)\\
&\quad=Vec(\mathbf{B}).
\end{aligned}
$$

Assuming that the needed matrix is invertible, one obtains the formula for Y,

$$
\begin{aligned}
Vec(Y)=&\left[\sum_{i,j}S_i^{-1/2}\widetilde{\omega}_j\bigotimes S_i^{-1/2}\widetilde{\omega}_j+\sum_i S_i^{-1/2}\bigotimes S_i^{-1/2}\right.\\
&\left.\quad-\sum_i S_i^{-1/2}\widetilde{\omega}_i\bigotimes S_i^{-1/2}-\sum_i S_i^{-1/2}\bigotimes S_i^{-1/2}\widetilde{\omega}_i\right]^{-1}Vec(\mathbf{B}).
\end{aligned}
\tag{16}
$$

Eq. (14) can be solved by using matrices of smaller size. Denote by $Vech(A)$ the $q(q+1)/2\times 1$ vector formed by stacking the subdiagonal elements of a $q\times q$ matrix A under each other, $Vech(A)=(a_{11},a_{21},\ldots,a_{q1},a_{22},\ldots,a_{2q},\ldots,a_{qq})^T$. Then for every symmetric matrix A, $Vec(A)=G_qVech(A)$, with the *duplicating* matrix G_q of size $q^2\times q(q+1)/2$ (see [5, Section 16.4] or [9, Section 3.8]). If H_q is a left inverse of G_q, (14) means that

$$
\begin{aligned}
&H_q\left[\sum_i S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)\bigotimes S_i^{-1/2}\,(\mathbf{I}-\widetilde{\omega}_i)+\sum_{i\neq j}S_i^{-1/2}\widetilde{\omega}_j\bigotimes S_i^{-1/2}\widetilde{\omega}_j\right]G_qVech(Y)\\
&\quad=Vech(\mathbf{B}),
\end{aligned}
$$

and $Vech(Y)$ can be obtained from this equation.

3.2. The Mandel–Paule procedure

When $q=1$, an easily implementable method for estimating the common mean was suggested by Mandel and Paule [10]. The goal here is to extend this algorithm to the vector situation.

Even without the normality assumption, (12) implies that for the optimal weights, $W_i=[Var(X_i)]^{-1}$,

$$
\begin{aligned}
&E\sum_i[Var(X_i)]^{-1/2}(X_i-\tilde{X})(X_i-\tilde{X})^T[Var(X_i)]^{-1/2}\\
&\quad=p\mathbf{I}-\sum_i[Var(X_i)]^{-1/2}\mathbf{W}^{-1}[Var(X_i)]^{-1/2}.
\end{aligned}
\tag{17}
$$

As in Section 3.1, the identity (17) can be used as the estimating equation for θ and Ξ, provided that Σ_i are estimated by S_i. The suggestion here is to choose Y, which is designed to estimate Ξ, as an (approximate) solution of the matrix equation,

$$\sum_i (S_i + Y)^{-1/2}(X_i - \tilde{X})(X_i - \tilde{X})^T (S_i + Y)^{-1/2}$$
$$+ \sum_i (S_i + Y)^{-1/2} \left[\sum_k (S_k + Y)^{-1} \right]^{-1} (S_i + Y)^{-1/2} = p\mathbf{I}. \tag{18}$$

Notice that (18) implies that

$$\mathbf{tr} \left(\sum_i (X_i - \tilde{X})^T (S_i + Y)^{-1} (X_i - \tilde{X}) \right) = (p-1)q,$$

and for $q = 1$ the second sum in the left-hand side of (18) reduces to 1. In this case the original Mandel–Paule algorithm recommends weights of the from (8) such that

$$\sum_i \frac{(X_i - \tilde{X})^2}{S_i + Y} = p - 1.$$

Eq. (18) can be interpreted as a simplified version of the restricted maximum likelihood equation. Indeed it follows from (6) that

$$(\hat{\Sigma}_i + Y)^{-1} = (\hat{\Sigma}_i + Y)^{-1} \left[(X_i - \hat{\theta})(X_i - \hat{\theta})^T + \left(\sum_k (\hat{\Sigma}_k + Y)^{-1} \right)^{-1} \right] (\hat{\Sigma}_i + Y)^{-1}$$
$$+ v_i \hat{\Sigma}_i^{-1} (\mathbf{I} - S_i \hat{\Sigma}_i^{-1}).$$

If $\hat{\Sigma}_k = S_k$, then

$$\sum_i (S_i + Y)^{-1/2}(X_i - \hat{\theta})(X_i - \hat{\theta})^T (S_i + Y)^{-1/2}$$
$$= p\mathbf{I} - \sum_i (S_i + Y)^{-1/2} \left[\sum_k (S_k + Y)^{-1} \right]^{-1} (S_i + Y)^{-1/2},$$

so that, if the estimates S_i from the individual studies are close to the REML estimators, then the REML estimator of Ξ must be close to the solution Y of (18). When $q = 1$, a similar result was derived in [12]. Asymptotic properties of the scalar version of the Mandel–Paule rule are investigated in [13].

To find an approximate solution, we look for solutions close to Y_{DL} and put $Y_{\mathrm{MP}} = Y_{\mathrm{DL}} + y$ with a "small" symmetric matrix y. Matrix differentiation [9, Section 8.1] shows that with

$$W_{\mathrm{DL}} = \sum_k (S_k + Y_{\mathrm{DL}})^{-1},$$

$$\mathbf{L} = \frac{1}{2} \sum_i (S_i + Y_{\mathrm{DL}})^{-1/2} [(X_i - \tilde{X}_{\mathrm{DL}})(X_i - \tilde{X}_{\mathrm{DL}})^T + W_{\mathrm{DL}}^{-1}]$$
$$\times (S_i + Y_{\mathrm{DL}})^{-1/2} \bigotimes (S_i + Y_{\mathrm{DL}})^{-1} + \frac{1}{2} \sum_i (S_i + Y_{\mathrm{DL}})^{-1} \bigotimes (S_i + Y_{\mathrm{DL}})^{-1/2}$$

$$
\begin{aligned}
&\times[(X_i-\tilde{X}_{\mathrm{DL}})(X_i-\tilde{X}_{\mathrm{DL}})^T+W_{\mathrm{DL}}^{-1}](S_i+Y_{\mathrm{DL}})^{-1/2}\\
&-\sum_{i,j}(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\\
&+\sum_{i,j}(S_i+Y_{\mathrm{DL}})^{-1/2}(X_i-\tilde{X}_{\mathrm{DL}})\tilde{X}_{\mathrm{DL}}^T W_{\mathrm{DL}}^{-1}\omega_j\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\\
&+\sum_{i,j}(S_j+Y_{\mathrm{DL}})^{-1/2}\omega_j\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}(X_i-\tilde{X}_{\mathrm{DL}})\tilde{X}_{\mathrm{DL}}^T W_{\mathrm{DL}}^{-1}\omega_j\\
&-\sum_{i,j}(S_i+Y_{\mathrm{DL}})^{-1/2}(X_i-\tilde{X}_{\mathrm{DL}})X_j^T(S_j+Y_{\mathrm{DL}})^{-1}\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\\
&-\sum_{i,j}(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}(X_i-\tilde{X}_{\mathrm{DL}})X_j^T(S_i+Y_{\mathrm{DL}})^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{R}=&\sum_i(S_i+Y_{\mathrm{DL}})^{-1/2}(X_i-\tilde{X}_{\mathrm{DL}})(X_i-\tilde{X}_{\mathrm{DL}})^T(S_i+Y_{\mathrm{DL}})^{-1/2}\\
&+\sum_i(S_i+Y_{\mathrm{DL}})^{-1/2}W_{\mathrm{DL}}^{-1}(S_i+Y_{\mathrm{DL}})^{-1/2}-p\mathbf{I},
\end{aligned}
$$

the vectorized version of (18) for the “correction” term y has the form,

$$\mathbf{L}\,Vec(y)=Vec(\mathbf{R}).$$

If the matrix $\mathbf{L}$ is non-singular,

$$Vec(y)=\mathbf{L}^{-1}\,Vec(\mathbf{R}). \tag{19}$$

As in Section 3.1, (19) can be replaced by

$$Vech(y)=[H_q\mathbf{L}G_q]^{-1}Vech(\mathbf{R}).$$

The simulations show that the solution obtained from the following simplified version $\mathbf{L}_{\mathrm{sim}}$ of the matrix $\mathbf{L}$ leads to a good numerical approximation to $Vech(y)$,

$$
\begin{aligned}
\mathbf{L}_{\mathrm{sim}}=&\frac{1}{2}\sum_i(S_i+Y_{\mathrm{DL}})^{-1/2}[(X_i-\tilde{X}_{\mathrm{DL}})(X_i-\tilde{X}_{\mathrm{DL}})^T+W_{\mathrm{DL}}^{-1}]\\
&\times(S_i+Y_{\mathrm{DL}})^{-1/2}\bigotimes(S_i+Y_{\mathrm{DL}})^{-1}+\frac{1}{2}\sum_i(S_i+Y_{\mathrm{DL}})^{-1}\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}\\
&\times[(X_i-\tilde{X}_{\mathrm{DL}})(X_i-\tilde{X}_{\mathrm{DL}})^T+W_{\mathrm{DL}}^{-1}](S_i+Y_{\mathrm{DL}})^{-1/2}\\
&-\sum_{i,j}(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j\bigotimes(S_i+Y_{\mathrm{DL}})^{-1/2}\omega_j.
\end{aligned}
$$

After y has been determined, one can take $Y_{\mathrm{MP}}=[Y_{\mathrm{DL}}+y]_+$ (so that Y_{MP} has the same spectral decomposition as $Y_{\mathrm{DL}}+y$, with the eigenvalues being positive parts of these of $Y_{\mathrm{DL}}+y$).

Thus, the Mandel–Paule rule is the weighted means statistic (7) whose matrix weights have the form,

$$W_i=(Y_{\mathrm{MP}}+S_i)^{-1}. \tag{20}$$

As the DerSimonian–Laird estimator, this extension of the Mandel–Paule rule provides the estimate $\tilde{X}$ of the common parameter θ along with the estimate Y_{MP} of Ξ.

We formulate our main results.

Proposition 3.1. *The DerSimonian–Laird estimator of the common vector mean θ is the weighted mean statistic* (9) *with the matrix weights* (8) *where $Y = Y_{\mathrm{DL}}$ solves* (14). *The vectorized form of this matrix satisfies* (16). *The Mandel–Paule rule is determined by the weights of the form* (8) *with Y found from* (18). *If the REML estimators of Σ_i are S_i, then the Mandel–Paule estimator is the REML estimator of Y. Its approximate version is given by $Y_{\mathrm{MP}} = Y_{\mathrm{DL}} + y$, where y satisfies* (19).

Now we turn to estimation of the covariance matrix of the weighted means statistics (7), as this estimator is needed for confidence ellipsoids.

4. Covariance matrix estimation

Here we discuss two estimators of the covariance matrix of the weighted means statistics (7) for arbitrary, but fixed, symmetric non-negative definite matrix weights W_i assuming that $\mathbf{W} = \sum_{i=1}^{p} W_i$ is a non-singular matrix. Denote $\omega_i = \mathbf{W}^{-1} W_i$, $i = 1, \ldots, p$. Then $\sum_{i=1}^{p} \omega_i = \mathbf{I}$, and all eigenvalues of ω_i are real positive numbers, smaller than one, although these matrices do not have to commute or to be symmetric or positive definite.

4.1. Unbiased quadratic procedures

We start with the search for a quadratic estimator in the residuals, $X_i - \tilde{X}$, of

$$Var(\tilde{X}) = \sum_{i=1}^{p} \omega_i Var(X_i) \omega_i^T .$$

In other terms the goal is to find $q \times q$ non-negative definite matrices Q_i, such that

$$\sum_{i=1}^{p} Q_i E(X_i - \tilde{X})(X_i - \tilde{X})^T Q_i^T = \sum_{i=1}^{p} \omega_i Var(X_i) \omega_i^T ,$$

or

$$\sum_{i=1}^{p} Q_i (\mathbf{I} - \omega_i) Var(X_i) (\mathbf{I} - \omega_i)^T Q_i^T + \sum_{i \neq j} Q_j \omega_i Var(X_i) \omega_i^T Q_j^T$$
$$= \sum_{i=1}^{p} \omega_i Var(X_i) \omega_i^T .$$

It follows that for any $i = 1, \ldots, p$,

$$Q_i (\mathbf{I} - \omega_i) Var(X_i) (\mathbf{I} - \omega_i)^T Q_i^T + \sum_{k:k \neq i} Q_k \omega_i Var(X_i) \omega_i^T Q_k^T = \omega_i Var(X_i) \omega_i^T .$$

As in Section 3.1, these equations hold for all symmetric matrices $Var(X_i)$ if and only if

$$H_q\left[Q_i(\mathbf{I}-\omega_i)\bigotimes Q_i(\mathbf{I}-\omega_i)+\sum_{k:k\neq i}Q_k\omega_i\bigotimes Q_k\omega_i\right]G_q$$
$$=H_q\left(\omega_i\bigotimes\omega_i\right)G_q. \tag{21}$$

The identity (21) shows that for $i=1,\ldots,p$,

$$H_q\left(Q_i\bigotimes Q_i\right)\left[(\mathbf{I}-\omega_i)\bigotimes(\mathbf{I}-\omega_i)-\omega_i\bigotimes\omega_i\right]G_q$$
$$+H_q\sum_k\left(Q_k\bigotimes Q_k\right)\left(\omega_i\bigotimes\omega_i\right)G_q=H_q\left(\omega_i\bigotimes\omega_i\right)G_q.$$

We rewrite these equalities in the form

$$H_q\left(Q_i\bigotimes Q_i\right)\left[(\mathbf{I}-\omega_i)\bigotimes(\mathbf{I}-\omega_i)-\omega_i\bigotimes\omega_i\right]G_qH_q$$
$$=H_q\omega_i\bigotimes\omega_iG_qH_q-H_q\sum_k\left(Q_k\bigotimes Q_k\right)\left[\omega_i\bigotimes\omega_i\right]G_qH_q,$$

and use the fact that for any $q\times q$ matrix A, $(A\bigotimes A)G_qH_q=G_qH_q(A\bigotimes A)$ [9, Section 3.8].

Assuming that matrices $(\mathbf{I}-\omega_i)\bigotimes(\mathbf{I}-\omega_i)-\omega_i\bigotimes\omega_i=\mathbf{I}\bigotimes\mathbf{I}-\omega_i\bigotimes\mathbf{I}-\mathbf{I}\bigotimes\omega_i$ are invertible, one gets

$$H_q\left(Q_i\bigotimes Q_i\right)G_qH_q$$
$$=\left[H_q\left(\omega_i\bigotimes\omega_i\right)G_qH_q-H_q\sum_k\left(Q_k\bigotimes Q_k\right)\left(\omega_i\bigotimes\omega_i\right)G_qH_q\right]$$
$$\times\left[(\mathbf{I}-\omega_i)\bigotimes(\mathbf{I}-\omega_i)-\omega_i\bigotimes\omega_i\right]^{-1}.$$

Let

$$\Omega_j=H_q\omega_j\bigotimes\omega_j\left[\mathbf{I}\bigotimes\mathbf{I}-\omega_j\bigotimes\mathbf{I}-\mathbf{I}\bigotimes\omega_j\right]^{-1}G_q. \tag{22}$$

Then multiplying by G_q from the left and summing up by i gives

$$\sum_i H_q\left(Q_i\bigotimes Q_i\right)G_q\left[\mathbf{I}+\sum_j\Omega_j\right]=\sum_j\Omega_j,$$

so that for $i=1,\ldots,p$,

$$H_q\left(Q_i\bigotimes Q_i\right)G_q=\Omega_i-\sum_j\Omega_j\left[\mathbf{I}+\sum_j\Omega_j\right]^{-1}\Omega_i=\left[\mathbf{I}+\sum_j\Omega_j\right]^{-1}\Omega_i. \tag{23}$$

Thus, an unbiased quadratic estimator of $Var(\tilde{X})$ exists if and only if the weights ω_i are such that the corresponding matrices (22) admit representation (23).

Notice that if $\lambda_1^{(i)}, \ldots, \lambda_q^{(i)}$ are the eigenvalues of ω_i, then Ω_i has eigenvalues $\lambda_s^{(i)}\lambda_t^{(i)}/(1-\lambda_s^{(i)}-\lambda_t^{(i)})$, $1 \leqslant s \leqslant t \leqslant q$. For example, if all matrices ω_i commute, then the eigenvalues $H_q(Q_i \bigotimes Q_i)G_q$ must be proportional to those of Ω_i, i.e.

$$\frac{\lambda_s^{(i)}\lambda_t^{(i)}}{1-\lambda_s^{(i)}-\lambda_t^{(i)}} = C\mu_s^{(i)}\mu_t^{(i)}$$

for $1 \leqslant s \leqslant t \leqslant q$. This can hold if and only if $\lambda_s^{(i)} \equiv \lambda^{(i)}$, i.e. if all ω_i are scalar multiples of $\mathbf{I}$, $\omega_i = \lambda^{(i)}\mathbf{I}$, $\sum_i \lambda^{(i)} = 1$, in which case when $\max_k \lambda^{(k)} < \frac{1}{2}$, $\Omega_i = \frac{(\lambda^{(i)})^2}{1-2\lambda^{(i)}}\mathbf{I}$, and

$$Q_i = \frac{\sqrt{\dfrac{(\lambda^{(i)})^2}{1-2\lambda^{(i)}}}}{\sqrt{1+\sum_k \dfrac{(\lambda^{(k)})^2}{1-2\lambda^{(k)}}}}\mathbf{I}.$$

In the case when $\lambda^{(i)} \equiv 1/p$, i.e. $\tilde{X}$ is the sample mean $\tilde{X}_\infty$, the unbiased quadratic estimator is the classical unbiased estimator,

$$\mathbf{S} = \frac{1}{p(p-1)}\sum_{i=1}^{p}(X_i-\tilde{X})(X_i-\tilde{X})^T.$$

If (23) holds, an estimator of $Var(\tilde{X})$ can be readily found. Indeed,

$$\begin{aligned} &Vech\left(\sum_{i=1}^{p} Q_i E(X_i-\tilde{X})(X_i-\tilde{X})^T Q_i^T\right) \\ &= \sum_{i=1}^{p} H_q\left(Q_i \bigotimes Q_i\right) G_q Vech(E(X_i-\tilde{X})(X_i-\tilde{X})^T) \\ &= \left[\mathbf{I}+\sum_j \Omega_j\right]^{-1}\left[\sum_i \Omega_i Vech(E(X_i-\tilde{X})(X_i-\tilde{X})^T)\right]. \end{aligned} \tag{24}$$

Thus, $Vech(\sum_{i=1}^{p}\omega_i Var(X_i)\omega_i^T)$ could be estimated by the right-hand side of (24). The same formula is amenable to calculation of the unbiased estimator of $Vech(Var(\tilde{X}))$ when more general estimators of the form

$$\sum_i C_i Vech((X_i-\tilde{X})(X_i-\tilde{X})^T)$$

with $q(q+1)/2 \times q(q+1)/2$ matrices C_i (not necessarily of the form (23)) are allowed. Of course such an estimate may not produce a positive definite matrix in which case its positive part is taken.

When $q = 1$, one obtains the formula originally derived in [2],

$$Q_i^2 = \frac{\omega_i^2}{1-2\omega_i}\left[1+\sum_k \frac{\omega_k^2}{1-2\omega_k}\right]^{-1}.$$

Obviously this is not a satisfactory solution if $\max_i \omega_i \geqslant \frac{1}{2}$.

4.2. *Almost unbiased estimators of a covariance matrix*

As in the general case an unbiased quadratic estimator of $Var(\tilde{X})$ does not exist, we put forward a different estimator which can be justified in a more general setting of linear models (such as (1)).

As before, let ω_i be fixed normalized matrix weights, $\sum_i \omega_i = \mathbf{I}$. To estimate the covariance matrix $Var(\tilde{X})$ of the (unbiased) weighted means statistic $\tilde{X}$, one can use the almost unbiased estimate of $Var(X_i)$ derived in [8] as follows. One has

$$\begin{aligned} Var(X_i - \tilde{X}) &= (\mathbf{I}-\omega_i)Var(X_i)(\mathbf{I}-\omega_i)^T + \sum_{k\neq i}\omega_k Var(X_k)\omega_k^T \\ &= \sum_k \omega_k Var(X_k)\omega_k^T + Var(X_i) - \omega_i Var(X_i) - Var(X_i)\omega_i^T. \end{aligned}$$

Of course, when $\omega_i = [\sum_{j=1}^p Var(X_j)^{-1}]^{-1} Var(X_i)^{-1}$, the first term in the right-hand side simplifies to

$$\sum_k \omega_k Var(X_k)\omega_k^T = \left[\sum_{i=1}^p Var(X_i)^{-1}\right]^{-1} = \frac{1}{2}\omega_i Var(X_i) + \frac{1}{2}Var(X_i)\omega_i^T,$$

and by substituting this expression in the previous formula, one obtains

$$Var(X_i - \tilde{X}) = Var(X_i) - \tfrac{1}{2}\omega_i Var(X_i) - \tfrac{1}{2}Var(X_i)\omega_i^T.$$

Horn et al. [8, p. 382] argue that by continuity if the weights are only approximately correct, this is an approximate identity. Thus, an *almost unbiased* estimator of $Var\left(X_i - \tilde{X}\right)$ is derived by solving (in V_i) the following equation

$$(X_i - \tilde{X})(X_i - \tilde{X})^T = V_i - \tfrac{1}{2}\omega_i V_i - \tfrac{1}{2}V_i\omega_i^T.$$

Here symmetric V_i serves as an estimate of $Var(X_i)$. This solution can be found like one in (23).

More precisely,

$$\begin{aligned} &Vech(\widehat{Var}(X_i)) \\ &\quad = \left[H_q\left(\mathbf{I} - \frac{1}{2}\omega_i \bigotimes \mathbf{I} - \frac{1}{2}\mathbf{I}\bigotimes \omega_i\right)G_q\right]^{-1} Vech((X_i - \tilde{X})(X_i - \tilde{X})^T). \end{aligned} \tag{25}$$

The same agreement as in Sections 2 and 3.1 about taking the positive part of a non-positively defined symmetric matrix is applicable here too. Actually, as in our situation, $Var(X_i) \geqslant \Sigma_i$ and an unbiased estimate S_i of Σ_i is available, it makes sense to use as the final estimator of $Var(X_i)$, $\max[\widehat{Var}(X_i), S_i] = S_i + [\widehat{Var}(X_i) - S_i]_+$ with $\widehat{Var}(X_i)$ determined from (25).

Table 1
The covariance matrices of seven estimators when $q = 2$; $p = 3, 6$, $n = 9$, 18 and their estimates

	$p=3, n=9$	$p=3, n=18$	$p=6, n=9$	$p=6, n=18$
$\bar{Y}$	0.68 0.07	0.67 0.05	0.35 0.04	0.33 0.02
	0.07 0.59	0.05 0.59	0.04 0.28	0.02 0.29
S	0.50 0.05	0.24 0.02	1.23 0.13	0.58 0.06
	0.05 0.44	0.02 0.22	0.13 1.08	0.06 0.51
GD	0.68 0.08	0.64 0.07	0.40 0.04	0.34 0.03
	0.08 0.64	0.07 0.57	0.04 0.35	0.03 0.31
(28)	0.08 0.01	0.10 0.00	0.03 0.00	0.04 0.01
	0.01 0.07	0.00 0.09	0.00 0.03	0.01 0.05
DL	0.54 0.06	0.53 0.07	0.26 0.03	0.25 0.02
	0.06 0.46	0.07 0.46	0.03 0.23	0.02 0.21
(26)	0.54 0.06	0.53 0.05	0.26 0.03	0.25 0.02
	0.06 0.48	0.05 0.48	0.03 0.23	0.02 0.22
(28)	0.59 0.07	0.58 0.05	0.31 0.03	0.29 0.03
	0.07 0.54	0.05 0.54	0.03 0.27	0.03 0.26
MP	0.55 0.08	0.54 0.07	0.28 0.03	0.27 0.02
	0.08 0.49	0.07 0.47	0.03 0.24	0.02 0.23
(26)	0.59 0.05	0.57 0.06	0.31 0.03	0.29 0.03
	0.05 0.53	0.06 0.52	0.03 0.27	0.03 0.26
(28)	0.42 0.05	0.43 0.05	0.21 0.02	0.22 0.02
	0.05 0.39	0.05 0.38	0.02 0.18	0.02 0.19
ML	0.54 0.04	0.53 0.08	0.27 0.04	0.24 0.02
	0.04 0.51	0.08 0.47	0.04 0.22	0.02 0.20
(5)	0.46 0.02	0.60 0.05	0.26 0.02	0.31 0.02
	0.02 0.31	0.05 0.55	0.05 0.23	0.02 0.28
R	0.54 0.03	0.53 0.02	0.26 0.02	0.24 0.02
	0.03 0.46	0.02 0.45	0.02 0.23	0.02 0.21
(26)	0.38 0.01	0.35 0.02	0.23 0.01	0.22 0.01
	0.01 0.32	0.02 0.31	0.01 0.21	0.01 0.21
(5)	0.25 0.00	0.31 0.02	0.15 0.01	0.20 0.01
	0.00 0.23	0.02 0.29	0.01 0.14	0.01 0.18
SN	0.59 0.05	0.54 0.08	0.30 0.04	0.25 0.03
	0.05 0.50	0.08 0.47	0.04 0.26	0.03 0.22
(26)	0.55 0.03	0.52 0.06	0.28 0.03	0.27 0.02
	0.03 0.44	0.06 0.46	0.03 0.24	0.02 0.23
(5)	0.36 0.00	0.49 0.05	0.21 0.02	0.25 0.02
	0.00 0.29	0.05 0.44	0.02 0.18	0.02 0.23

The resulting formula for $Var(\tilde{X})$ estimator has the form,

$$\widehat{Var}(\tilde{X}) = \sum_i \omega_i \widehat{Var}(X_i)\omega_i^T, \tag{26}$$

which can be recovered from its vectorized version via (25),

$$Vech(\widehat{Var}(\tilde{X})) = \sum_i G_q \left(\omega_i \bigotimes \omega_i\right) H_q \times \left[H_q \left(\mathbf{I} - \frac{1}{2}\omega_i \bigotimes \mathbf{I} - \frac{1}{2}\mathbf{I} \bigotimes \omega_i\right) G_q\right]^{-1} Vech((X_i - \tilde{X})(X_i - \tilde{X})^T). \tag{27}$$

The statistic (26) gives an estimate of the covariance matrix of any weighted means statistic for fixed weights ω_i. This estimate is uniquely defined when $\mathbf{I} - \frac{1}{2}\omega_i \bigotimes \mathbf{I} - \frac{1}{2}\mathbf{I} \bigotimes \omega_i$ is invertible, which holds provided that $\mathbf{I} - \omega_i$ is non-singular.

When $\omega_i \equiv p^{-1}\mathbf{I}$, i.e. when $\tilde{X}$ is the sample mean, this estimator coincides with the classical unbiased estimator $\mathbf{S}$.

Proposition 4.1. *An unbiased quadratic estimator*, $\sum_{i=1}^{p} Q_i(X_i - \tilde{X})(X_i - \tilde{X})^T Q_i^T$, *of the covariance matrix* $\sum_{i=1}^{p} \omega_i Var(X_i)\omega_i^T$ *for fixed weights* ω_i *is defined when the matrices* Ω_i *given by* (22) *admit representation* (23), *in which case its vectorized form is given by the right-hand side of* (24). *The almost unbiased estimator* (26) *of* $Var(\tilde{X})$ *satisfies* (25) *and its vectorized version is given by* (27).

An alternative estimator of $Var(\tilde{X})$ is

$$\widetilde{Var}(\tilde{X}) = \left[\sum_i \widehat{Var}(X_i)^{-1}\right]^{-1}. \tag{28}$$

In particular, this estimate is commonly used for the maximum likelihood estimator although it is known to underestimate the true variance. Exact theoretical comparison of (28) and (26) is difficult, but simulations reported in the next section suggest that (26) is a better estimate of this variance.

5. Example and simulation results

Here we report the results of a Monte-Carlo simulation study when $q = 2$ or 3, $p = 3$ or 6. The distribution of S_i was taken to be that of a Wishart random matrix with parameters $n_i - 1$ and Σ_i. The distribution of the within-laboratory covariance matrices, Σ_i, was the inverse Wishart distribution with parameters $q + 2$ and $\mathbf{I}$, so that $E\Sigma_i = \mathbf{I}$. We chose $\Xi = [1, 0.2; 0.2, 0.8]$ for $q = 2$ and $\Xi = [1.8, 0.2, 0.1; 0.2, 2.1, 0.4; 0.1, 0.4, 2.5]$ when $q = 3$. The sample sizes vector either had equal coordinates, $n_i \equiv n = 9$; $n = 18$, or n was a permutation of integers $1, \ldots, p$ plus 5.

Table 2
The estimators of $\Xi = [1.00, 0.20; 0.20, 0.80]$, $q = 2$ when $p = 3, 6, n = 9, 18$

	$p = 3, n = 9$	$p = 3, n = 18$	$p = 6, n = 9$	$p = 6, n = 18$
y_{DL}	1.23 0.19	1.19 0.19	1.11 0.20	1.06 0.19
	0.19 1.05	0.19 1.03	0.20 0.93	0.19 0.86
y_{MP}	0.91 0.14	0.91 0.14	0.93 0.16	0.95 0.16
	0.14 0.80	0.14 0.77	0.16 0.77	0.16 0.78
$\hat{\Xi}$	0.56 0.03	0.67 0.10	0.63 0.10	0.83 0.14
	0.03 0.38	0.10 0.48	0.10 0.53	0.14 0.68
$\hat{\Xi}_{SN}$	0.70 0.08	0.71 0.12	0.98 0.10	0.95 0.11
	0.08 0.64	0.12 0.58	0.10 0.88	0.11 0.83
$\hat{\Xi}_{R}$	0.44 0.05	0.63 0.10	0.67 0.11	0.85 0.13
	0.05 0.38	0.10 0.55	0.11 0.58	0.13 0.70

Table 3
The estimators of $\Xi = [1.0, 0.2, 0.1; 0.2, 1.2, 0.8; 0.1, 0.8, 1.6]$, $q = 3$ when $p = 3, 6$; n is a permutation of $(1, \ldots, p)+5$

	$p = 3$			$p = 6$		
y_{DL}	1.20	0.15	0.07	1.13	0.17	0.11
	0.15	1.36	0.75	0.17	1.35	0.73
	0.07	0.75	1.80	0.11	0.73	1.93
y_{MP}	0.90	0.13	0.09	0.77	0.13	0.06
	0.13	1.04	0.62	0.13	1.00	0.54
	0.09	0.62	1.28	0.06	0.54	1.41
$\hat{\Xi}$	0.89	0.12	0.09	0.94	0.15	0.10
	0.12	1.11	0.48	0.15	1.02	0.12
	0.09	0.48	1.23	0.10	0.12	1.09
$\hat{\Xi}_R$	0.94	0.11	0.05	1.48	0.09	0.06
	0.11	1.08	0.41	0.09	1.63	0.38
	0.05	0.41	1.20	0.06	0.38	1.84
$\hat{\Xi}_{SN}$	1.02	0.14	0.02	1.09	0.16	0.01
	0.14	1.10	0.06	0.16	1.15	0.19
	0.02	0.06	1.40	0.01	0.19	1.45

Table 4
The covariance matrices of six estimators for q, p and n as in Table 3

	$p = 3$			$p = 6$		
DL	0.25	0.01	0.00	0.23	0.01	0.04
	0.01	0.26	0.10	0.01	0.27	0.10
	0.00	0.10	0.33	0.04	0.10	0.45
(26)	0.24	0.01	0.01	0.22	0.02	0.01
	0.01	0.27	0.10	0.02	0.26	0.09
	0.01	0.10	0.33	0.01	0.09	0.42
MP	0.28	0.01	0.00	0.24	0.01	0.05
	0.01	0.30	0.09	0.01	0.29	0.09
	0.00	0.09	0.36	0.05	0.09	0.51
(26)	0.31	0.02	0.00	0.27	0.03	0.05
	0.02	0.32	0.10	0.03	0.33	0.11
	0.00	0.10	0.41	0.05	0.11	0.54
ML	0.27	0.03	0.01	0.25	0.01	0.04
	0.03	0.26	0.10	0.01	0.24	0.10
	0.01	0.10	0.30	0.04	0.10	0.40
(26)	0.31	0.02	0.00	0.27	0.03	0.05
	0.02	0.32	0.10	0.03	0.33	0.11
	0.00	0.10	0.41	0.05	0.11	0.54
R	0.28	0.03	0.01	0.25	0.01	0.04
	0.03	0.25	0.13	0.01	0.24	0.10
	0.01	0.13	0.33	0.04	0.10	0.40
(26)	0.31	0.02	0.00	0.27	0.03	0.05
	0.02	0.32	0.10	0.03	0.33	0.11
	0.00	0.10	0.41	0.05	0.11	0.54
SN	0.27	0.01	0.00	0.25	0.01	0.04
	0.01	0.25	0.10	0.01	0.24	0.10
	0.00	0.10	0.30	0.04	0.10	0.40
(26)	0.31	0.02	0.00	0.27	0.03	0.05
	0.02	0.32	0.10	0.03	0.33	0.11
	0.00	0.10	0.41	0.05	0.11	0.54

Table 1 contains the covariance matrices of the studied estimators. The DerSimonian–Laird method, and the Mandel–Paule algorithm (which behaved very similarly to the DerSimonian–Laird method) systematically were among the best procedures. They did remarkably well compared to the "golden standard" of maximum likelihood procedures which are more computationally intensive (Tables 2, 3). The Graybill–Deal estimator *GD* has a variance matrix substantially exceeding that of all other estimators, and the sample mean was the second worst. For this reason we excluded these estimators in Table 4 which gives the simulated values of the covariance matrices of six estimators and their estimates via (26), (28) or (5) when $q = 3$ and the sample sizes of $p = 3, 6$ laboratories are chosen at random. Clearly, (26) gives a much better estimate of the covariance matrix of the weighted means procedures than (28). The latter is especially inadequate in the case of the Graybill–Deal estimator *GD*. Because of heterogeneity the sample covariance matrix **S** turns out to be a rather poor estimator of the covariance matrix of the sample mean.

Tables 2 and 3 contain the estimators of the between-laboratory effect Ξ for these values of q, p and n. They show that the likelihood estimators systematically underestimate Ξ while the Mandel–Paule procedure, Y_{MP}, and the DerSimonian–Laird method, Y_{DL}, gave better answers. However, since the DerSimonian–Laird rule is computationally simpler than the Mandel–Paule algorithm, we can recommend it in the multivariate case (and it does not need any recommendations in the scalar case being the tool-of-choice for meta-analysis.)

Our simulations also demonstrated that the distribution of the pivotal ratio, $(\hat{\theta}-\theta)^T[\widehat{Var}(\hat{\theta})]^{-1}(\hat{\theta}-\theta)$ can be well approximated by a multiple of F-distribution with q and $p-q$ degrees of freedom (Fig. 1).

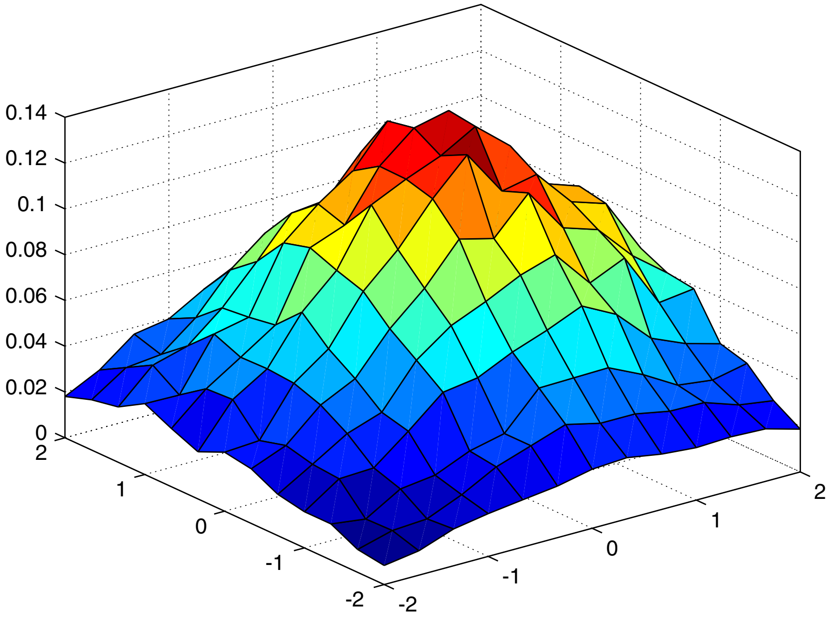

Fig. 1. The plot of the estimated density of $[\widehat{Var}(\hat{\theta})]^{-1/2}(\hat{\theta}-\theta)$ for the maximum likelihood estimator when $q = 2$, $p = 7$, and $\Xi = [1, 0.2; 0.2, 0.8]$.

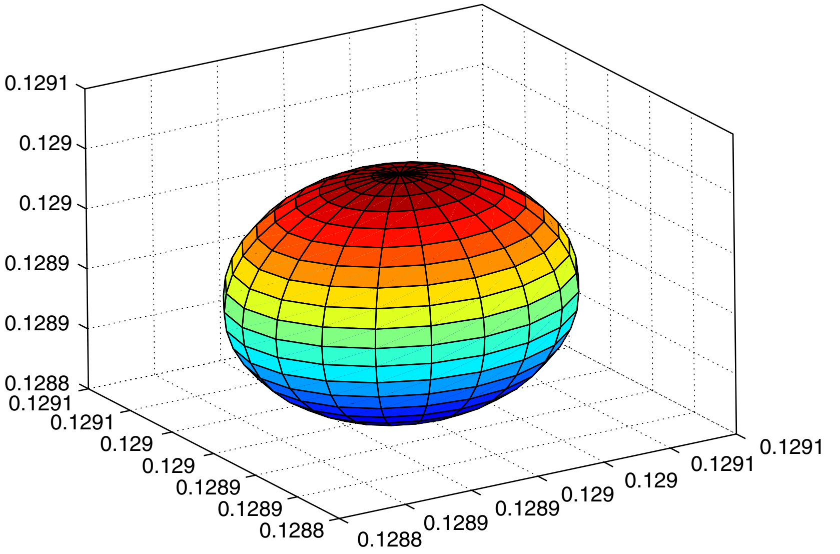

Fig. 2. Confidence ellipsoid for the charge sensitivity in the Key Comparisons of accelerometers, $q = 3$, $p = 14$, Y_{DL} is the zero matrix.

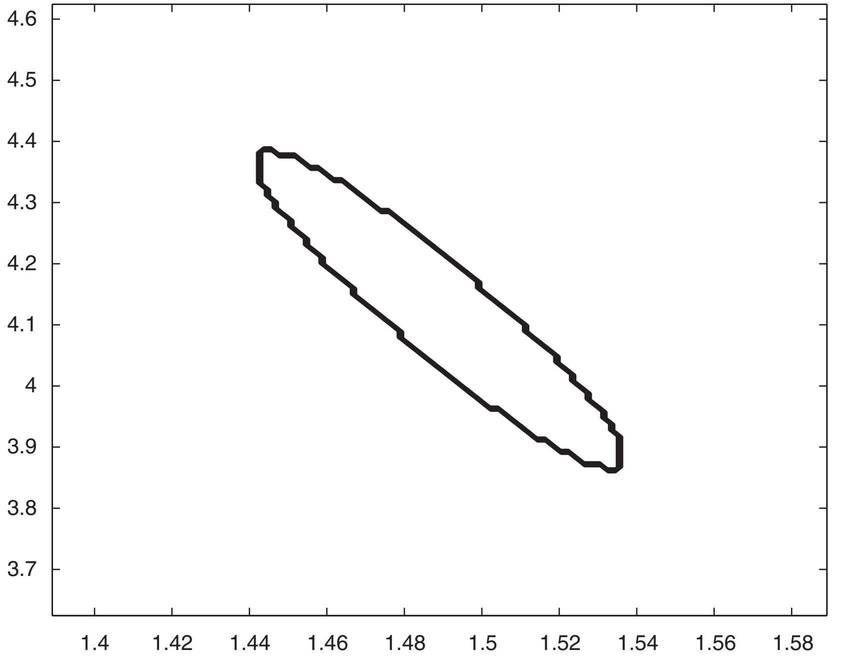

Fig. 3. Confidence ellipse for the diffusivity and thermal heat in the thermal diffusivity example, $q = 2$, $p = 28$, and $Y_{DL} = [0.084, -0.277; -0.277, 0.914]$.

To illustrate techniques of Section 3 we implemented them in the accelerometers key comparisons study [11] mentioned in Section 1 and in interlaboratory study of thermal diffusivity and conductivity. Only the results for the frequencies 40, 50 and 63 Hz are given here (Fig. 2).

The Mandel–Paule procedure and the DerSimonian–Laird method gave virtually the same answer for $\theta^T = [0.12678, 0.12667, 0.12672]$, with the zero matrix $\hat{\Xi}$ and the estimated covariance matrix $\widehat{Var}(\tilde{X}) = 1.0e - 07*$ $[0.3982, 0.0222, -0.0206 : 0.0222, 0.3849, 0.0075; -0.0206, 0.0075, 0.3856]$.

This matrix leads to an approximate confidence ellipsoid for θ on the basis of F-distribution with $q = 3$ and $p - q = 11$ degrees of freedom. This ellipsoid provides useful information about the joint behavior of the charge sensitivities for these frequencies. Notice that the maximum likelihood estimators were slightly different: $[0.12675, 0.12671, 0.12667]$ (Fig. 3).

References

[1] W.G. Cochran, The combination of estimates from different experiments, Biometrics 10 (1954) 101–129.

[2] N. Cressie, Playing safe with mis-weighted means, J. Amer. Statist. Assoc. 77 (1982) 597–754.

[3] R. DerSimonian, N. Laird, Meta-analysis in clinical trials, Controlled Clinical Trials 7 (1986) 177–188.

[4] F.A. Graybill, R.B. Deal, Combining unbiased estimators, Biometrics 15 (1959) 543–550.

[5] D.A. Harville, Maximum likelihood approaches to variance components estimation and related problems, J. Amer. Statist. Assoc. 72 (1977) 320–340.

[6] D.A. Harville, Matrix Algebra from a Statistician's Perspective, Springer, New York, 1997.

[7] L.V. Hedges, I. Olkin, Statistical Methods for Meta-Analysis, Academic Press, Orlando, FL, 1985.

[8] R.A. Horn, S.A. Horn, D.B. Duncan, Estimating heteroscedastic variance in linear models, J. Amer. Statist. Assoc. 70 (1975) 380–385.

[9] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, New York, 1999.

[10] J. Mandel, R.C. Paule, Interlaboratory evaluation of a material with unequal number of replicates, Anal. Chem. 42 (1970) 1194–1197.

[11] H.J. von Martens, C. Elster, A. Link, A. Taebner, W. Wabinski, Final Report on Key Comparison CCAUV.V-K1, Metrologia Technical Supplement, vol. 40, 2003. ⟨www.iop.org/EJ/abstract/0026-1394/40/1A/09001⟩

[12] A.L. Rukhin, B. Biggerstaff, M.G. Vangel, Restricted maximum likelihood estimation of a common mean and Mandel–Paule algorithm, J. Statist. Plann. Inference 83 (2000) 319–330.

[13] A.L. Rukhin, M.G. Vangel, Estimation of a common mean and weighted means statistics, J. Amer. Statist. Assoc. 93 (1998) 303–308.

[14] S. Searle, G. Casella, C. McCulloch, Variance Components, Wiley, New York, 1992.

[15] M.G. Vangel, A.L. Rukhin, Maximum likelihood analysis for heteroscedastic one-way random effects ANOVA in interlaboratory studies, Biometrics 55 (1999) 129–136.